CentER

Discussion paper

No. 2001-69

ON D-OPTIMALITY BASED TRUST REGIONS FOR BLACK-BOX OPTIMIZATION PROBLEMS

By Lonneke Driessen, Ruud Brekelmans, Herbert Hamers and Dick den Hertog

September 2001

ISSN 0924-7815

On D-optimality Based Trust Regions for Black-box Optimization Problems

Lonneke Driessen * Ruud Brekelmans † Herbert Hamers ‡

Dick den Hertog ‡

September 7, 2001

Abstract

In this paper we show how techniques from response surface methodology and mathematical programming can be combined into a new sequential derivative-free approach for solving unconstrained deterministic black-box optimization problems. In this sequential derivative-free optimization approach local approximations of the underlying objective function are optimized within a trust region framework. If the points that determine the local approximations are located in such a way that the approximations become bad, a geometry improving iteration is carried out instead of an objective improving iteration.

We incorporate the D-optimality criterion, well-known in design of experiments, in our approach in two different ways. Firstly, it is used to define a trust region that adapts its shape to the locations of the points in which the objective function has been evaluated. Secondly, it determines an optimal geometry improving point. An attractive feature of our approach is that it is insensitive to affine transformations.

Keywords: D-optimality, trust region, derivative free, optimization, affine transformations.

1 Introduction

Black-box optimization problems are common in, for example, design optimization, where time-consuming function evaluations are often carried out by simulation tools. Due to freedom in design, there usually are a huge number of design alternatives. However, as the simulation of one design already takes quite some time, only a fraction of the total number of possible designs can be evaluated. To cope with this complexity, techniques from statistics and mathematical optimization like design of experiments (DoE), Response Surface Methodology

*Corresponding author, Centre for Quantitative Methods BV, P.O. Box 414, 5600 AK Eindhoven, The Netherlands, E-mail: Driessen@cqm.nl

†Tilburg University, CentER Applied Research

‡Tilburg University, Department of Econometrics and Operations Research

(RSM), and derivative free optimization, are more and more recognized as being indispensable in this area.

The unconstrained deterministic black-box optimization problem we consider, is stated formally as

$$\min_{d} \; f(d) \qquad d \in \mathbb{R}^q, \tag{P$_0$}$$

where q denotes the number of design parameters. The analytical form of the objective function f is unknown, as well as any derivative information. Hence, the only way to get information about this function is by evaluating it for distinct design points, or in short points, d. In this paper we assume that each function evaluation is expensive or time-consuming.

Response Surface Methodology (RSM) (e.g., Myers and Montgomery (1995)), is frequently used in industry for process and product design and optimization. RSM is a collection of statistical design techniques, empirical model building, and optimization methods. In RSM the response surface function is approximated by linear and quadratic models. The algorithm iterates through the following sequence. The approximating models are optimized, the found optimum is simulated, and with help of this new simulation point the approximating models are updated. Hence, improvements are achieved by performing additional experiments along the path of steepest descent (in case of a minimization problem). Basically, this boils down to optimizing a local linear approximation under a spherical trust region constraint.

Also in the field of mathematical programming methods have been developed to solve unconstrained black-box optimization problems. Alexandrov et al. (1998) present a framework for generating a sequence of approximations to an expensive objective function based on pattern search methods. The derivative free methods by both Conn and Toint (1996) and Powell (2000) iteratively form quadratic models by interpolation and optimize these models in a trust region framework. The trust region takes the shape of a sphere around the best solution found so far. Hence, it does not take the positions of the simulated points into account. Furthermore, the trust region depends on the scaling of the design parameters. An interesting aspect of these methods is that the points that determine the local approximations should satisfy some geometric conditions (e.g., Powell (1994a), Conn and Toint (1996), Conn et al. (1997), Powell (2000)). If these geometric conditions are not satisfied, a geometry improving step is carried out. Some iterations of the method in Powell (1994a) are aimed at modifying the shape of some simplex, in order that interpolation at its vertices is likely to yield good linear models of objective and constraints. Conn and Toint (1996) follow Powell (1994b) by using the determinant of the design matrix as a geometry measure.

Summarized, both RSM and mathematical programming techniques solve the unconstrained black-box optimization problem (P$_0$) iteratively by optimizing local low-order polynomial approximations within a trust region. As the methods do not use derivative information, the local models are fitted on a set of points already simulated. In each iteration of the algorithm it is decided which of the simulated points are included in the set of points on which the local models are fitted. The trust region has the shape of a sphere and is centered around the best point found so far. A disadvantage of this type of trust region is that its location and shape are chosen independently of the location of the points on which the local models are fitted. Hence, if the design points are located in a long, narrow area, the trust region is still spherical, while it would be more accurate to apply a trust region with an adapted shape in this situation. Another drawback of both the RSM and mathematical programming techniques is that they are sensitive to an affine transformation of the input variables. More precisely, the solution to the transformed problem and the transformation of the solution to the original problem are not necessarily the same. This implicates that, dependent on the chosen scaling, the algorithm evolves in a different way.

We suggest a new trust region that accounts for the disadvantages explained above. It automatically incorporates the information about the location of the points and is insensitive to affine transformations. This is achieved by linking the classical approach used in RSM and mathematical programming to the D-optimality criterion for finding good DoE schemes. This leads to the use of ellipsoidal trust regions. The position of the center of the ellipsoid as well as its shape and rotation depend on the location of the points on which the local models are fitted. The D-optimality criterion is also used in the geometry improving step of the optimization.

The rest of this paper is organized as follows. In Section 2 we introduce our new ellipsoidal trust region. Instead of a spherical trust region like the one used by Powell (1994a), we use an on D-optimality based ellipsoidal trust region. Section 3 proposes a new method for geometry improvement, also based on the D-optimality criterion. In Section 4 we show that use of our ellipsoidal trust region results in an optimization method that is insensitive to affine transformations of the design space under consideration. Some conclusions and future research are outlined in Section 5.

2 The ellipsoidal trust region

We follow the approach to solve the black-box optimization problem (P_0) by iteratively optimizing local linear approximations within a trust region. In this section we first formulate the classical trust region model. Then we give the intuition behind our new ellipsoidal trust region and formulate the resulting new model.

In both the steepest descent approach in RSM and the Mathematical Programming approach followed by Powell (2000) and Conn and Toint (1996) the new evaluation point is determined by optimizing the approximating model under a trust region constraint. We focus on linear approximations instead of the quadratic approximations used by Powell (2000) and Conn and Toint (1996). The classical problem can then be stated as[1]

$$\begin{aligned} &\max_d \ \beta' d \\ &\text{s.t.} \left\| \, d - d^* \right\|_2 \leq \Delta, \end{aligned} \qquad (\mathrm{P}_1)$$

where $\beta \in \mathbb{R}^q$ is the vector of model coefficients arising, for example, from a linear least squares regression on the design points, $d \in \mathbb{R}^q$ is the decision

[1]For ease of exposition the constant term has not been included in this formulation of the optimization model. Its presence does not alter the optimal solution of (P_1).

variable, the vector of design parameters, $d^* \in \mathbb{R}^q$ is the best point found so far, and $\Delta \in \mathbb{R}$ is the trust region radius. Hence, a local linear model of the underlying real model is optimized within a spherical trust region.

The location of the points on which the local approximating models are fitted does not influence the shape of this spherical trust region in any way. Furthermore, in this trust region framework it is implicitly assumed that all design parameters are of comparable magnitude. The dispersion of the design points contains information about the reliability of the models fitted on these points. In order to be able to incorporate this information into the method we formulate a new problem by changing the shape of the trust region. In statistics, a natural way to take locations of design points into account is by using the prediction variance of the approximation. Stochastic models have been used for deterministic simulations before, see Sachs et al. (1989). As long as the variance remains within acceptable ranges, the model is trusted. The idea is to apply this approach to our deterministic problem. This is possible by assuming that there is some prediction variance present. We will show that this variance is minimized in the center of gravity of the design points and that the contour curves of this variance are ellipsoids. Before formulating this result formally in Theorem 1, we first introduce some notation. The results of Theorem 1 allow us to choose a suitable trust region during the optimization process.

In the statistic linear regression model the variance of the predictor $\hat{y} = x'\beta_{+0}$ equals $\sigma_e x'(X'X)^{-1}x$. The matrix X is known as the extended design matrix and consists of the row vectors $(x^i)' = [1\ (d^i)'], i = 1, \ldots, n$, where d^i denotes the design vector for the i^{th} experiment and n is the number of design points. The matrix X is assumed to have linearly independent columns. With β_{+0} we denote the model coefficients including the coefficient for the constant term and σ_e denotes the variance of the normally distributed error. Although the usual statistical conditions on the error term are not satisfied in the deterministic situation, we assume that σ_e equals 1. The covariance matrix $(X'X)^{-1}$ plays an essential role in D-optimality. We will show how this matrix induces a new trust region. As we focus on the design space and do not take the constant term into account, we work with the matrix C instead of the matrix $(X'X)^{-1}$, which is related to $(X'X)^{-1}$ in the following way

$$(X'X)^{-1} = \begin{pmatrix} a & b' \\ b & C \end{pmatrix}, \tag{1}$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}^q$, and $C \in \mathbb{R}^{q \times q}$. As $(X'X)^{-1}$ is positive definite and symmetric, C is positive definite and symmetric as well. Hence, C is also nonsingular. Due to the special structure of the matrix X, the matrix $X'X$ has the following structure:

$$X'X = \begin{pmatrix} n & n\bar{d} \\ n\bar{d} & D'D \end{pmatrix}, \tag{2}$$

where $\bar{d}$, the center of gravity of the design points $d^i, i = 1, ..., n$, is defined as

$$\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d^i$$

and the matrix D equals X without the first column of ones, i.e., $X = [1\ D]$.

Theorem 1 *The variance of the predictor* $\hat{y} = X\beta_{+0}$ *is minimal in* $\bar{d}$, *the center of gravity of the design points* $d^i, i = 1, ..., n$. *The contour curves of this variance are given by the ellipsoids*

$$(d-\bar{d})'C(d-\bar{d}) = \rho, \tag{3}$$

where $\rho = \rho_0 - a + \bar{d}'C\bar{d}$ *and* ρ_0 *equals the variance of the predictor.*

Proof: Kleijnen et al. (2001) showed that the variance is minimal in the point $-C^{-1}b$. By substituting equations (1) and (2) into the identity $(X'X)^{-1}(X'X) = I$ and then decompose it, we find that

$$\begin{pmatrix} na + nb'\bar{d} & na\bar{d}' + b'D'D \\ nb + nC\bar{d} & nb\bar{d}' + CD'D \end{pmatrix} = \begin{pmatrix} 1 & 0' \\ 0 & I_{q\times q} \end{pmatrix}. \tag{4}$$

By comparing the two block entries on position (2, 1) of this identity we find that

$$\bar{d} = -C^{-1}b, \tag{5}$$

which proves the first part of the proposition.

The second part can be proved by recalling that the variance of the predictor in a certain design point equals $x'(X'X)^{-1}x$. Using equation (1) and the identity $x' = [1\ d']$, we find that

$$\begin{aligned} x'(X'X)^{-1}x &= d'Cd + 2b'd + a \\ &= d'Cd - 2\bar{d}'Cd + a \\ &= (d-\bar{d})'C(d-\bar{d}) - \bar{d}'C\bar{d} + a. \end{aligned} \tag{6}$$

Since ρ_0 equals the variance of the predictor, we have

$$x'(X'X)^{-1}x = \rho_0 \tag{7}$$

This equation denotes a contour curve for the variance of the predictor. Combining (6) and (7) it follows immediately that (3) holds. □

Theorem 1 shows that the ellipsoids arising from the matrix C are in fact contour curves of the variance. We propose to use the ellipsoids based on the matrix C in the definition of the trust region. The new problem formulation including the ellipsoidal trust region becomes

$$\begin{aligned} &\max_d \beta' d \\ &\text{s.t.} \parallel d - \bar{d} \parallel_C \leq \rho, \end{aligned} \tag{P$_2$}$$

where ρ is the trust region radius and $\|x\|_C$ is the C-norm defined by

$$\|x\|_C = \sqrt{x'Cx}.$$

The first of the two main differences between problem (P_2) and problem (P_1) is that we now use the C-norm instead of the 2-norm. As the matrix C is positive definite, it defines a proper matrix norm. The second difference between problem (P_2) and problem (P_1) is that in problem (P_2) the center of the trust region is determined by all the design points on which the local linear models are fitted together, while in problem (P_1) the trust region is centered around the best point so far.

We illustrate the implications of using this C-norm instead of the 2-norm in Figure 1. Two important observations arise from this example. The first one is that the ellipsoidal trust region adapts its form to the locations of the design points, whereas the spherical trust region does not. This adaptation ensures that the models are more trusted in areas where actual evaluations have been performed. The second observation is that the center of the ellipsoidal trust region is determined by the design points such that the ellipsoid covers the design points in the best possible way. The spherical region is centered around the best point found so far. Hence, if such a point lies a bit apart from the other design points, some parts of the spherical trust region might not contain design points at all.

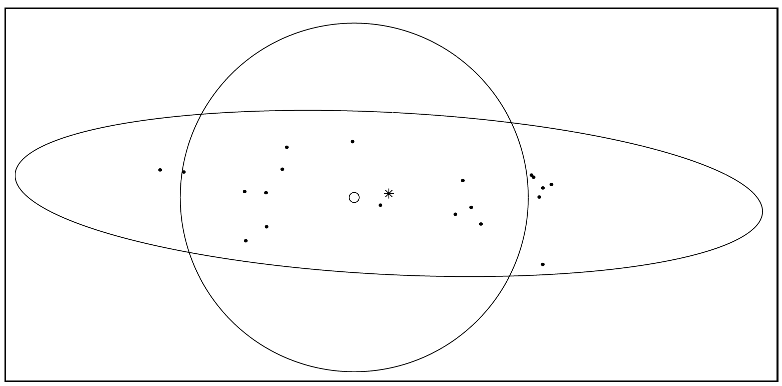

Figure 1: *The ellipsoidal trust region adapts better to the locations of the design points than the spherical trust region. The small black dots indicate design points, the $*$ indicates the center of the ellipse, and the open dot is the center of the sphere.*

When the ellipsoidal trust region becomes too narrow in one or more directions, this is an indication that the consecutively simulated design points have more or less the same value for these dimensions. Eventually, the approximating models will start to show lack of fit in these dimensions. In the next section we describe how to prevent the occurrence of this situation.

3 Geometry improvements

It is desirable to develop an algorithm that ensures that the optimization process will not get stuck because of a bad positioning of the design points on which the local models are based. Otherwise, the quality of the fitted models might become very poor and wrong conclusions are drawn. This problem is dealt with by incorporating a geometry check in the algorithm. When this check points out that the geometry of the design points is poor, a geometry improving simulation is carried out instead of an objective improving simulation.

We describe how to incorporate the ideas behind the new ellipsoidal trust region also in the geometry improving iterations of the optimization process. We discuss the difference with the method used by Powell (1994b) and Conn et al. (1997). Finally, we show the correspondence between our geometry improving step and the D-optimality criterion in statistical DoE.

Powell (1994b) proposes to concentrate on the determinant of the extended design matrix, $\det(X)$. He uses interpolation to find local approximations and therefore the extended design matrix X is always square in his method. Conn et al. (1997) also apply this approach. The mathematical program that must be solved to find the geometry improving point is

$$\begin{aligned} &\max_{d} \ \det(X(d)) \\ &\text{s.t.} \left\| \, d - d^* \right\|_2 \leq \Delta, \end{aligned} \tag{P$_3$}$$

where $X(d)$ denotes the extended design matrix after inclusion of design point d. We use this notation throughout the rest of this paper to denote a matrix after inclusion of the design point between brackets. Powell reasons that the determinant of a square matrix is a measure for the degree of singularity of this matrix. It is desirable to work with a non-singular extended design matrix as it is used for solving a linear system of equations to create the interpolation models. Golub and van Loan (1996) (p. 82) though, point out that matrices with a low absolute value for the determinant exist that are far from singular, as well as matrices with a high absolute value of the determinant that are almost singular. Hence, in certain situations the determinant of a square matrix is not a good measure for the degree of singularity of this matrix.

A first way to implement a geometry improving step is to use the D-optimality criterion. A second possibility is to use our trust region matrix C. The first approach is inspired by the commonly used D-optimality criterion from DoE (see Myers and Montgomery (1995)). In DoE the problem of extending an existing design in the best possible way is a well-known problem. By intuition, a geometry improving step is performed when the locations of the design points are such that some dimensions of the design space are hardly explored. By performing a geometry improving step we wish to maximize the amount of information that can be obtained from the design. This is exactly what the D-optimality criterion is about. A design of experiments is D-optimal when the generalized variance, $\det(X'X)^{-1}$, is minimized. This minimization is desired because the hyper volume of the joint confidence region of the $\beta's$ is proportional to $\sqrt{\det(X'X)^{-1}}$. Not only the volume, but also the ellipsoidal shape of the confidence region depends on $X'X$. Hence, the criterion based on D-optimality is to minimize $\det((X'X)^{-1}(d))$, i.e., to find the d that, when added to the extended design matrix X, leads to a minimal value for $\det(X'X)^{-1}$.

The second approach is based on our trust region matrix C. It is logical to aim at maximizing the volume of the trust region in (P$_2$) in a geometry improving step. For a fixed value of ρ maximizing the volume of our ellipsoid is equivalent with maximizing the determinant of the matrix C^{-1}, as the volume of the ellipsoid is given by

$$\mu_q \sqrt{\det(\rho^2 C^{-1})},$$

where μ_q denotes the volume of the q-dimensional unit sphere, which depends only on q. Hence, the second criterion is to maximize $\det(C(d)^{-1})$ or equivalently, to minimize $\det(C(d))$, i.e., to find the d such that, when added to the design matrix D, leads to a minimal value for $\det(C)$.

These two different approaches lead to two possible objectives in the geometry improving step, minimize $\det((X'X)^{-1}(d))$ and minimize $\det(C(d))$. In the following theorem we prove that these two objectives lead to the same optimal solution.

Theorem 2 *The matrix C, defined in (1), satisfies:*

$$\det(C) = n \det(X'X)^{-1}.$$

Proof: We recall that $X'X$ has a special structure (see (2)). The following relation holds for the matrix $X'X$ (Schur complement):

$$\begin{pmatrix} 1 & 0 \\ -\bar{d} & I \end{pmatrix} \begin{pmatrix} n & n\bar{d}' \\ n\bar{d} & D'D \end{pmatrix} = \begin{pmatrix} n & n\bar{d}' \\ 0 & D'D - n\bar{d}\bar{d}' \end{pmatrix}.$$

Hence

$$\det(X'X) = n \det(D'D - n\bar{d}\bar{d}')$$

and

$$\det(X'X)^{-1} = \frac{1}{n \det(D'D - n\bar{d}\bar{d}')}. \tag{8}$$

From the block entries (2, 2) in (4) combined with (5) it follows that

$$C^{-1} = D'D - n\bar{d}\bar{d}'.$$

Hence

$$\det C = \frac{1}{\det(D'D - n\bar{d}\bar{d}')}. \tag{9}$$

The proposition follows by combining (8) and (9). □

We conclude that minimization of $\det C$ is a good geometry improving objective, both from a theoretical as well as from an intuitive point of view: in DoE a lot of research has been done in D-optimal designs and they have proved to work well, and intuitively it is appealing to maximize the volume of the trust region. The geometry objective used by Powell (1994b) and the one we derived here, are equivalent. Hence, with a different reasoning we return to the measure proposed by Powell.

Besides the objective function we also have to constrain the area in which the best geometry improving design should be located. Without such a region constraint, the optimal design point would be located as far as possible away from the other design points. Of course, a design point too far away is not useful anymore for fitting *local* approximating models. Our first idea, to use the ellipsoidal trust region here as well, is not applicable. A bad geometry means that there are some dimensions of the design space that are not enough explored compared to others. The shape of the ellipsoidal trust region reflects this bad positioning by a small range for the relatively unexplored dimensions and a large range for the other dimensions. Figure 2 illustrates the situation. This implicates that, when using the ellipsoidal trust region, the search region for the geometry improving point is very narrow in the dimensions we are most interested in to explore more.

Another disadvantage of using the ellipsoidal trust region in the geometry improving step is the following. Dykstra (1971) showed that all the points on a certain ellipsoid have the same generalized variance, $\det(X'X)^{-1}$. As our objective in the geometry improving step is to minimize this generalized variance, any point on the ellipsoidal trust region constraint would be optimal.

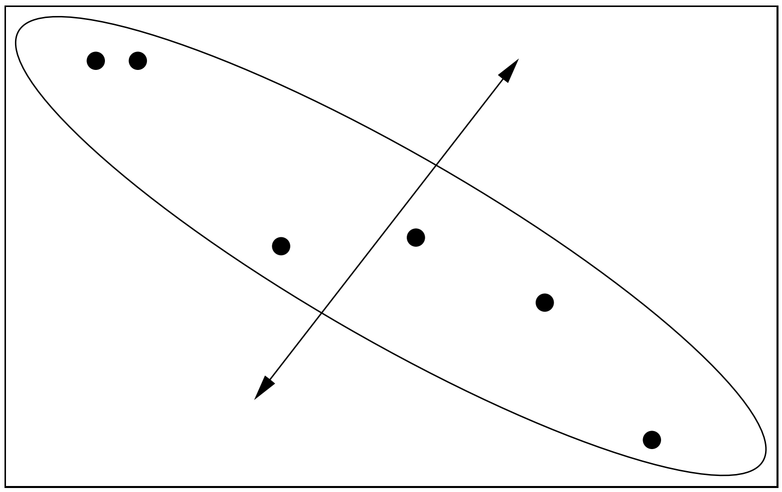

Figure 2: *The direction in which exploration is most desirable for geometry improvement is also most restricted by the ellipsoidal trust region.*

If the design problem is scaled in such a way that all dimensions are of equal magnitude, the classical spherical trust region would be most appropriate as trust region for the geometry improving step. Often though, the design problem is not nicely scaled already and we propose here to incorporate the matter of scaling in the trust region constraint in the following way. We assume that the transformation $\xi(d) = Ad$ turns the design problem into a properly scaled problem. Here, A is a user-provided square non-singular matrix of dimension q. We propose to apply the following trust region for the geometry improving step

$$\| d - \bar{d} \|_{(A'A)} \leq \tau,$$

where τ denotes the radius of the area in which the optimal geometry improving point should be searched for. Note that also this trust region is an ellipsoid. It does not adapt to the locations of the design points though. It only incorporates the scaling effects. To avoid confusion with the on variance based trust region for the objective improving step and the on design scaling based trust region for the geometry improving step, from now on we name the latter search region instead of trust region. In Section 4 we show that use of this search region results in a geometry improving step that is insensitive to affine transformations on the design space.

The mathematical program that has to be solved to find the geometry improving point is

$$\begin{aligned} &\min_{d} \ \det C(d) \\ &\text{s.t.} \parallel d - \bar{d} \parallel_{(A'A)} \leq \tau. \end{aligned} \tag{P$_4$}$$

Dykstra (1971) has shown the surprising fact that the objective function in problem (P_4) is equivalent to a quadratic objective function. The problem then becomes a quadratic programming problem in which we wish to maximize a non-concave quadratic form over an ellipsoid. For this class of problems efficient solution methods exist, see for example Ye (1992).

4 Properties of the ellipsoidal trust region and search region

In this section we prove that use of our ellipsoidal trust and search regions makes the method insensitive to affine transformations of the design space. First we show that use of the classical spherical trust region implies an optimization method that is sensitive to such transformations.

The spherical trust region constraint

In the classical approach in each objective improving step problem (P_1) is solved. We show that the classical method including the spherical trust region is sensitive to affine transformations. Suppose that we transform the original problem by $\phi(d)$. Then we consider a method to be insensitive to affine transformations when the optimal solution of the original problem, d^*, and the optimal solution of the transformed problem, $\tilde{d^*}$ are related by $\tilde{d^*} = \phi(d^*)$. From now on, to distinguish between the variable space before and after the transformation we use the tilde sign above transformed variables.

Problem (P_1) is sensitive to a linear transformation of the variables d to $\tilde{d}$, defined by $\phi(d) = Md - s$, where M is a square, non-singular matrix of dimension $q \times q$ and s is a q-dimensional vector. In the rest of this paper M and s retain this meaning. If M is diagonal, pre-multiplication by it actually results in a scaling of the individual design parameters. Again we define $\tilde{x}' = [1 \ \tilde{d}']$. Then it follows that the linear transformation ϕ results in $\tilde{x}' = [1 \ (Md - s)']$. Hence, the constant term remains unaltered. Note that we can express the transformation ϕ in terms of x by pre-multiplying x by a matrix V with the following form

$$V = \begin{pmatrix} 1 & 0 \\ -s & M \end{pmatrix}, \tag{10}$$

where M and s are as defined above.

We see that problem (P_1) is sensitive to a linear transformation of the variables defined by ϕ by analyzing what happens to the variables x when multiplying them with V. The extended design matrix $\tilde{X}$ then becomes

$$\tilde{X} = XV'.$$

Substituting this into the normal equations for $\tilde{\beta}_{+0}$, i.e., $\tilde{\beta}_{+0} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y$, we find that[2]

$$\tilde{\beta}_{+0} = V^{-T}\beta_{+0},$$

from which follows that for the design space

$$\tilde{\beta} = M^{-T}\beta.$$

Hence, the transformed problem of (P_1)

$$\begin{aligned} &\max_{\tilde{d}} \ \tilde{\beta}'\tilde{d} \\ &\text{s.t. } \|\tilde{d} - \tilde{d}^*\|_2 \leq \Delta, \end{aligned}$$

can be rewritten to

$$\begin{aligned} &\max_{d} \ \beta' M^{-1} M d \\ &\text{s.t. } \|M(d - d^*)\|_2 \leq \Delta. \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\max_{d} \ \&\beta' d \\ &\qquad \text{s.t. } \|M(d - d^*)\|_2 \leq \Delta. \end{aligned}$$

We see that *in general* the problem is *not* insensitive to a transformation with ϕ. The trust region constraint is transformed with M, while the objective did not change. This means that we search an optimal solution to the same objective function over a different region now. We conclude that in general the objective improving step in the problem with a spherical trust region constraint is not insensitive to an affine transformation with ϕ. Note that this is caused by the fact that, though the objective is insensitive to such transformations, the trust region is not. We remark that for the special case of translation of d with a vector $s \in \mathbb{R}^q$, i.e. $M = I$, both d and d^* are translated with the same vector s and hence the feasible region does not change.

Next we concentrate on the geometry improving step for the problem with a spherical trust region constraint. This Problem (P_3) is not insensitive to a linear transformation of the variables defined by $\phi(d) = Md - s$. To see this we analyze how the solution of the with ϕ transformed problem relates to the original problem. After transformation the problem to solve becomes

$$\begin{aligned} &\max_{d} \ \det\left(X(d)V'\right) \\ &\text{s.t. } \|M(d - d^*)\|_2 \leq \Delta. \end{aligned}$$

[2] For ease of exposition we use x^{-T} to denote the inverse transposed of x.

As V is constant, this is equivalent to

$$\max_d \det(X(d))$$
$$\text{s.t. } \|M(d-d^*)\|_2 \leq \Delta.$$

The objective function is not influenced by the transformation, but when we look at the constraint, we see that the trust region constraint now contains the matrix M. This means that we search an optimal solution over a different region, just like in the objective improving step. Hence, in both steps of the classical optimization algorithm the objective function is insensitive to affine transformations, while the spherical trust region is not. The latter causes the optimization method as a whole to be sensitive to affine transformations.

The ellipsoidal trust region and search region constraints

In the following paragraphs we take a closer look at the impact of using an ellipsoidal trust and search region on the sensitivity to affine transformations. We first show that our objective improving step, solving problem (P_2), is insensitive to affine transformations on the design space.

Theorem 3 *Problem (P_2) is insensitive to a linear transformation of the variables defined by $\phi(d) = Md - s$.*

Proof: The only difference between problem (P_2) and the classical problem (P_1) is the fact that in the trust region constraint the C-norm is used instead of the 2-norm. We have shown already for problem (P_1) that the objective function is not influenced by a transformation by ϕ. We now take a closer look at the trust region constraint

$$\| \tilde{d} - \tilde{d}^* \|_{\tilde{C}} \leq \rho.$$

As $\tilde{d} = Md - s$, we can rewrite the left-hand side of this equation to

$$\| M(d-d^*) \|_{\tilde{C}} = (d-d^*)' M' \tilde{C} M (d-d^*).$$

Hence, if we can prove that $\tilde{C} = M^{-T} C M^{-1}$, then the trust region constraint is invariant under the transformation, which means that problem (P_2) is insensitive to the linear transformation ϕ.

We have seen already that the extended design matrix after transformation, $\tilde{X}$, equals XV'. Hence, the expression for $(\tilde{X}'\tilde{X})^{-1}$ becomes

$$(\tilde{X}'\tilde{X})^{-1} = V^{-T}(X'X)^{-1}V^{-1}.$$

With help of (1) and (10) we can rewrite (4) to

$$\begin{pmatrix} \tilde{a} & \tilde{b}' \\ \tilde{b} & \tilde{C} \end{pmatrix} = \begin{pmatrix} 1 & s'M^{-T} \\ 0 & M^{-T} \end{pmatrix} \begin{pmatrix} a & b' \\ b & C \end{pmatrix} \begin{pmatrix} 1 & 0 \\ M^{-1}s & M^{-1} \end{pmatrix}.$$

Writing out this block multiplication and concentrating on block element (2, 2) of the resulting right-hand side matrix, we find that

$$\tilde{C} = M^{-T} C M^{-1}.$$

$\square$

Next we concentrate on the geometry improving step, described by the mathematical program (P_4). We show that our geometry improving step, solving problem (P_4), is insensitive to affine transformations on the design space.

Theorem 4 *Problem (P_4) is insensitive to a linear transformation of the variables defined by $\phi(d) = Md - s$.*

Proof: The objective of problem (P_4), minimize $\det(C)$, is equivalent to maximizing $\det(X'X)$ according to Theorem 2. The extended design matrix after transformation, $\tilde{X}$, equals XV'. The translated objective then becomes minimization of $\det^2(V)\det(X'X)$ as V is a non-singular matrix. As V is constant, the solution to this minimization problem equals the solution to the original minimization problem before transformation. Hence, the objective of problem (P_4) is independent of transformations of the form ϕ.

Next we look at the search region constraint

$$\| \, d - \bar{d} \, \|_{(A'A)} \leq \tau.$$

Since A is the scaling matrix in the original design space, AM^{-1} is the right scaling matrix for the transformed situation. Hence,

$$\tilde{A} = AM^{-1}.$$

The left-hand side of the transformed search region constraint becomes

$$\begin{aligned}
\left\| \, \tilde{d} - \bar{\tilde{d}} \, \right\|_{(\tilde{A}'\tilde{A})} &= \| \, M(d - \bar{d}) \, \|_{((AM^{-1})'(AM^{-1}))} \\
&= \sqrt{(d - \bar{d})'M'((AM^{-1})'(AM^{-1}))M(d - \bar{d})} \\
&= \sqrt{(d - \bar{d})'A'A(d - \bar{d})} \\
&= \| \, d - \bar{d} \, \|_{(A'A)} \, .
\end{aligned}$$

Hence, the search region of the transformed problem equals the search region for the original problem. $\square$

5 Conclusions and future research

In this paper we propose the use of an ellipsoidal trust region constraint based on statistical D-optimality in a sequential optimization method for solving black-box optimization problems. The most attractive feature of this trust region is the fact that its shape and center are dependent on the location of the design points on which the approximating models are based. The D-optimality criterion is also used as geometry improvement objective. We showed the intuition behind the incorporation of the D-optimality criterion. Furthermore, we proved the independency of affine transformations for the ellipsoidal trust region.

Now that we have worked out our ideas in theory, our next step will be to implement this ellipsoidal trust region applied to unconstrained problems in our sequential design optimization toolbox SEQUEM (see Brekelmans et al. (2001)). If the results are promising we will thereafter incorporate the ellipsoidal trust region in the toolbox and adapt the method to constrained problems as well. Some of the interesting questions still remaining are how to deal with weighted regression and how to incorporate more complex models than the linear models we concentrated on in this paper.

Acknowledgement

We gratefully acknowledge the financial support from EIT-io. We also thank Jack Kleijnen, Arjen Vestjens and Peter Stehouwer for their useful suggestions and comments.

References

Alexandrov, N., J.E. Dennis, J., Lewis, R., and Torczon, V. (1998). A trust region framework for managing the use of approximation models in optimization. *Structural Optimization*, 15(1):16–23.

Brekelmans, R., Driessen, L., Hamers, H., and den Hertog, D. (2001). Constrained optimization involving expensive function evaluations: a sequential approach. Working paper.

Conn, A. and Toint, P. (1996). An algorithm using quadratic interpolation for unconstrained derivative free optimization. In *Nonlinear Optimization and Applications*, pages 27–47. Plenum.

Conn, A. R., Scheinberg, K., and Toint, P. L. (1997). Recent progress in unconstrained nonlinear optimization without derivatives. *Mathematical Programming*, 79:379–414.

Dykstra, O. (1971). The augmentation of experimental data to maximize $|X'X|$. *Technometrics*, 13:682–688.

Golub, G. and van Loan, F. (1996). *Matrix computations.* The John Hopkins University Press, 3^{rd} edition.

Kleijnen, J., den Hertog, D., and Angun, E. (2001). Searching for the maximum output in random simulation: New signal/noise heuristics. Working paper.

Myers, R. and Montgomery, D. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* John Wiley & Sons, New York.

Powell, M. (1994a). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis, Proceedings of the Sixth Workshopon Optimization and Numerical Analysis, Oaxaca, Mexico*, volume 275, pages 51–67, Dordrecht, The Netherlands. Kluwer Academic Publishers.

Powell, M. (1994b). A direct search optimization method that models the objective by quadratic interpolation. In *Presentation at the 5th Stockholm Optimization Days*.

Powell, M. (2000). UOBYQA: Unconstrained optimization by quadratic approximation. Numerical Analysis Report DAMTP 2000/NA14, University of Cambridge.

Sachs, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experimenets. *Statistical Science*, 4:409–435.

Ye, Y. (1992). A new complexity result on minimization of a quadratic function over a sphere constraint. In Floudas, C. and Pardalos, P., editors, *Recent advances in global optimization*, pages 19–21. Princeton University Press.